SAMGrid Developments - Part II

SAMGrid Project Status Report
18 February 2004

.

.

Robert D. Kennedy
Computing Division / CDFCA

Outline

$\Rightarrow$ **Continuing where Sinisa's talk (Part I) left off...**

$\Rightarrow$ **Completed and On-going Projects**

Grid-Jim Developments
Data Dimensions Grammar (COMPLETED)
Chains and Links
CD DFC Migration
CDF (H Stream) Reprocessing
Linux Deployment of Centralized Services
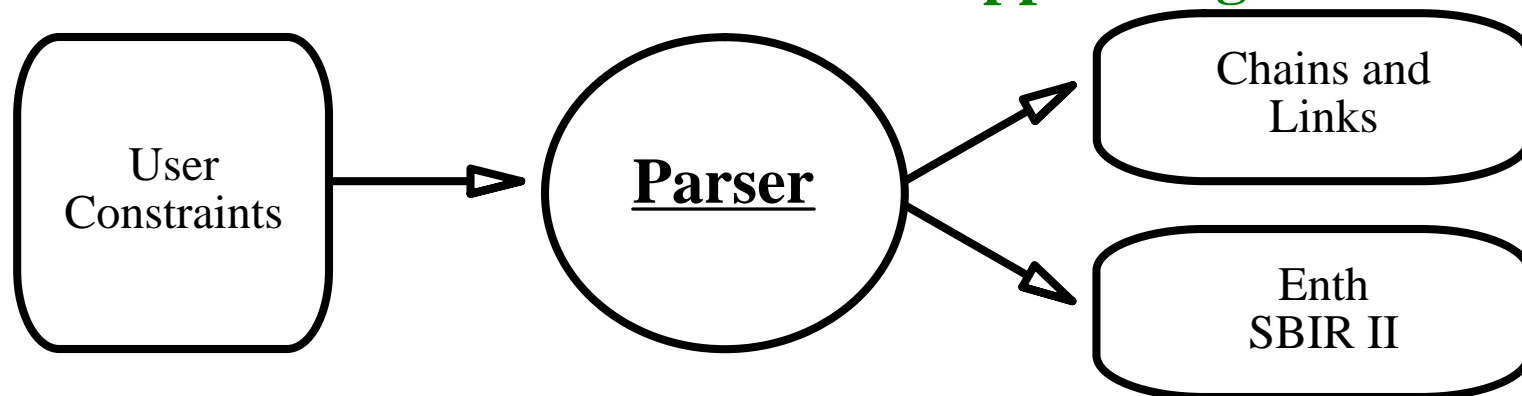**Station Developments... to become DH and Cache Services**
Sub-Project Assessments

$\Rightarrow$ **Segue to the Future....**

Grid-Jim Developments

⇒ **Mission: From http://www-d0.fnal.gov/computing/grid**
"... enhancing SAM... incorporating standard Grid tools and protocols"

⇒ **Main areas of work: Job Handling, Monitoring, GridFtp.**

⇒ **Who: Igor Terekhov, Gabriele Garzoglio, 2 students.**

⇒ **Deliverables (examples): Remote Job Submission (JIM), Run-Time Environment tarball management via SAM, Output datafile merging (DAGMAN), GridFtp and related authentication support.**

⇒ **Milestones: D0 Simulation Production (early March)**

⇒ **Status: Basic operations demonstrated, without merging**

⇒ **Follow-ons: Improved brokering, CDF JIMdeployment (June)**

Data Dimensions Grammar

⇒ **Mission:** Define/parse/support new "language" for specifying constraints for defining datasets, improve syntax checking.

⇒ **Who:** Randy Herber and Carmenita Moore.

⇒ **Deliverables:** Well-defined grammar for that "language", parser for that grammar, large test suite of observed queries (from existing mechanism) to test this all against.

⇒ **Status: Completed.**

⇒ **Follow-ons:** User-accessible doc of supported grammar.

User Constraints → **Parser** → Chains and Links / Enth SBIR II

Chains and Links

⇒ **Mission:** Extend "language" for specifying constraints for defining datasets. One of two impl.s which "compiles" parser output to SQL, performs queries, delivers results.

⇒ **Who:** Randy Herber, Carmenita Moore, Anil Kumar, Julie Trumbo.

⇒ **Tasks Remaining:** Chains and Links Table Schema, SAM db_server code modification, Loads of Chains and Links Tables with join information, Loads of existing dimensions information into new tables, SQL Builder (stand-alone app), Documentation

⇒ **Status:** Suffering from over-subscription of key personnel w/higher priority sub-projects. Schedule is being revised.

⇒ **Follow-ons:** Extend current functionality (more general results)

CDF DFC Migration

⇒ **Mission:** Replace CDF DFC schema and meta-data with SAM's schema and meta-data for all DH tasks.

⇒ **Who:** Randy Herber, Rick St. Denis.

⇒ **Deliverables:** Translator - DFC to SAM meta-data (done),
Modify CDF code to store meta-data directly into SAM,
Replace existing DFC-specific tools w/SAMtools,
If possible, maintain user interface - change impl to SAM.

⇒ **Milestones:** Completion June 01, 2004 (translator turned off)

⇒ **Status:** Being revived after hiatus to bring SAM to full production status at CDF and to reduce cost of maintaining redundant meta-data systems. Top priority for Randy's time.

⇒ **Note:** For CDF: some non-DH bits of DFC survive this.

CDF (H Stream) Reprocessing

⇒ **Mission:** Demonstrate CDF reprocessing large dataset on remote compute farm using SAM and CAF, with output files being stored back into SAM. (H Stream is large...)

⇒ **Who:** Frank Wuerthwein, Stefan Stonjek (SAM liaison).

⇒ **Deliverables:** Complete reprocessing scripts (divide "job" into manageably-sized projects), perform the reprocessing, store the data back into SAM.

⇒ **Milestones:** Finish scripts, start reprocessing by April 15.

⇒ **Status:** Sub-project in infancy. Doc and planning in progress.

⇒ **Notes:** Some overlaps with Grid-Jim's D0 Sim Prod. Main SAM issue is the processing/handling of VERY large datasets.

Linux - Centralized Services

⇒ **Mission:** To deploy SAMGrid centralized services on commodity Linux platforms instead of existing Sun platforms. And document the deployment procedures.

⇒ **Who:** Art Kreymer, Stefan Stonjek, Steve White, Dennis Box

⇒ **Milestone 1:** SAM db_server and CORBA name service in production on March 1 at CDF. Wait on web services.

⇒ **Milestone 2:** SAM web server ready for use by April 1 at CDF.

⇒ **Status:** Hardware is in place at FCC, burning in.

⇒ **Follow-ons:** Ensure high availability through (a) protocol specific mechanisms (eg CORBA IORs with fail-overs) or (b) generic mechanisms (Linux High Availability Project design with on-the-fly IP address fail-over).

Station Developments

⇒ **Mission:** Maintain and evolve the SAM "station", the SAM Data Handling and Cache management component.

⇒ **Who:** Andrew Baranovski

⇒ **Deliverables:** Increased flexibility, driven by field experience. <u>Many</u> improvements to support ever more diverse use-cases. External Storage Mechanism (ESM) to allow other storage elements (HPSS, dCache) to be treated like local cache.

⇒ **Milestones:** Take a 2 week freeze on station development, in order to merge current and db_server branches by March 1

⇒ **Follow-ons:** After the merge, we propose to restructure this project into another: DH and Cache Services (next page), and recaste this work into evolving sub-project model.

DH and Cache Services

⇒ **Mission:** Consider re-alignment of SAM "station" with GRID services model, modularize it into separable components.

⇒ **Who:** Andrew Baranovski, Rob Kennedy.

⇒ **Applications:** Improve interchangeability and cooperation of different "caches" (next version of ESM) using SRM, improve maintainability of "station" via modularization, prepare for GRID-era re-usability of components, consider applicability of GRID web services model.

⇒ **Milestones:** First SAM Design presentation March 11

⇒ **Status:** Re-org accepted, but sub-project not yet fully defined.

⇒ **Example Now:** In the dCache ESM, SAM does not know what is in cache, so no optimization of file delivery order. Usable, but we want better for global SAM deployment.

Sub-Project Assessments

⇒ **Mission:** Produce a uniform 2-page informal assessment of each SAM sub-project, sign-off by all directly involved.

⇒ **Who:** Rob Kennedy and Sinisa Veseli (lead on each varies).

⇒ **Procedure:** Informal and minimize impact on participants.
  (a) we ask for current docs to read and arrange interview,
  (b) 1 hour interview to fill out template with participants,
  (c) we write up outcome, obtain consensus by e-mail.

⇒ **Milestones:** Complete all first round assessments by March 1.

⇒ **Status:** All but one interview done. Drafts for nearly all out.

⇒ **Follow-ons:** Each new SAM sub-project will require an initial assessment to document its mission, deliverables, milestones. Each active sub-project will updates theirs every 1-3 months. Lead on these will pass from Rob/Sinisa to SPLs over time.

Assessment Template v1.1

⇒ **Background Documents**

⇒ **Project definition, mission station, organization (who)**

⇒ **Deliverables and time-table**

⇒ **Inter-project Dependencies**

⇒ **Project Status**

⇒ **Challenges and Critical Path Items**

⇒ **Lessons Learned**

⇒ **Project-specific Comments, alternative points-of-view, etc.**

⇒ **Next Sub-Project Assessment**

Segue to the Future...

⇒ **Much development accomplished in the past year, even more in the pipeline, some high-impact deployments soon.**

⇒ **We are positioning ourselves to manage a growing, diversifying list of development sub-projects.**
    - assessments provide uniform snapshot of state of SAMGrid
    - evolving towards hierarchical "flavored" organization
    - expanded mgmt team for more internal/external interactions

⇒ **More varieties of project relationships to consider as well.**
    - view before was mostly "internal" and "external"
    - also have "associations" with experiment projects,
    - soon "partnerships" with other major projects, ...

⇒ **SAMGrid development in the future: towards interchangeable GRID components while maintaining operations.**